AUDITORY INTERFACES: THE HUMAN PERCEIVER

H. Steven Colburn
Boston University
Boston, Massachusetts

P- 2.

This talk is a brief introduction to the basic auditory abilities of the human perceiver with particular attention toward issues that may be important for the design of auditory interfaces. The importance of appropriate auditory inputs to observers with normal hearing is probably related to the role of hearing as an omnidirectional, early warning system and to its role as the primary vehicle for communication of strong personal feelings.

Several basic properties of the human auditory perceiver should be kept in mind when designing interfaces. The range and resolution of the perceiver is impressive in terms of frequency (three decades with sensitivity to changes of 0.2%) and intensity (110 decibels with sensitivity to changes of a fraction of a decibel at almost all levels) so that the standards for clean sounds are quite high. On the other hand, our memory for absolute stimuli is much more limited, particularly with respect to intensity. In an identification paradigm, perceptual systems are generally limited to about 7 ± 2 intensity levels, so that the ability to identify is much more limited than the ability to differentiate. The notion of the critical bandwidth is also important in understanding auditory phenomena and abilities. In its crudest form, the critical band notion, which is consistent with a wide variety of data, is that the auditory system behaves as if it contained peripheral filters with bandwidths of about one-tenth of an octave.

The issues that arise in the design of auditory interfaces are different according to whether the interface is an isomorphic representation of the acoustic environment with natural sounds and trans-formations or a non-isomorphic mapping from fundamentally different inputs. In the completely isomorphic case, the problem is conceptually straightforward but very difficult to realize. Although one simply has to recreate the appropriate acoustic stimuli (from a robot or from a simulation) at the ears of the human listener, there are substantial engineering challenges in reproducing or creating the appropriate sounds. In addition to the obvious requirements for excellent fidelity in the acoustic sys-tem, the echo and reverberation effects on each sound must be appropriate and, critically, the motion of the listener's head must be coupled to the robot or program. The complexity of the problem can be illustrated by thinking of the effect on the sounds received at the ears by several sound sources in a normal environment. In this case, as the head moves, the "head-related transfer function" of each source changes in a way that can only be specified with knowledge of the position of the source and the position and orientation of the head. The processing of the human perceiver imposes important constraints on the quality of the reproduced sound signal.

In the simplest non-isomorphic cases, such as a scaled robot or a normal-sized robot in an envi-ronment with acoustic properties different than air (e.g., a helium rich environment), there are differ-ent problems, including the translations of signals to signals appropriate for human listeners. For example, a very small robot would have very small interaural differences that could be essentially undetectable to the human perceiver with an associated loss of ability to judge the azimuth of a

sound source. If these signals are processed to provide an appropriate stimulus on a human scale, the influence of the environment on the signals would have to be transformed and this would require significant knowledge of an environment that may be unfamiliar to the robot or its controller. The transformation of the received stimuli to provide appropriate acoustic inputs for a natural (human sized) situation requires a relatively full knowledge of the sound field.

In the case of more extreme deviations from a humanoid robot or simulation, the issues become conceptually difficult to think about. For example, if the robot has multiple ears so that a more sophisticated picture of the acoustic environment could be computed by the interface, the mapping of the signals or the information to the two ears of the human perceiver is difficult to design. Either some of the information will be neglected or we must learn how to increase the presentation of information to the human, possible by parsing the acoustic field and presenting multiple acoustic objects. This is a higher level recoding problem, and depends on techniques that have not yet been developed. Some of the attributes that result in the perception of separate acoustic objects are known and include separate spatial location, consistent modulation envelopes in frequency and amplitude, consistent harmonic structure, and others.

In some cases, if a stimulus is to be perceived as being generated by the listener, then there are special problems due to the fact that the sound is being received over two pathways, by the air-conduction pathway and by the bone-conduction pathway. The air-conducted sound can be monitored and transformed appropriately by the interface, but the bone-conduction pathway is difficult to cancel so that the stimulus perceived by the subject is a combination of the bone-conducted sound and the air-conducted sound provided by the interface. Learning would be a problem here if the new environment was trying to simulate an environment with light or heavy gases for example. In this case, the auditory stimulus would be a combination of the higher-pitched sound and the normal-pitched sound.

There are many cases in which unnatural stimuli are more effective than naturally occurring stimuli. [The argument that the human sensori-motor system is optimized by evolution (often applied to speech signals) cannot be applied in many cases because the constraints of the system are so complex and even difficult to know. In addition to questions of acoustics, for example, one can ask how important is the eating function to the design of the mouth or the necessity of being born to the size of the head?] When there is a well-defined task to be performed, it is often advantageous to provide non-natural processing of the stimuli. For example, angle resolution near the midline can be improved significantly by presenting nonlinearly processed stimuli that would increase the interaural differences even though some distortion would be added that could be important for some stimuli. On the other hand, when the tasks are extremely varied or unpredictable, there are significant advantages to using natural stimuli and allowing the experience of the listener to be used to primary advantage with little training required for many tasks. In addition, there are situations in which an unnatural stimulus may be advantageous such as the design of hearing aids with expanded ranges of levels and frequencies or in the design of auditory displays for non-auditory information.